

Please note

The text in this file has been automatically extracted and may contain minor errors. For the original version please consult the paper copy held in the Swinburne Library.

CHAPTER 1

INTRODUCTION

The increasing needs of an information based **society** has imposed tremendous pressures on our telecommunication system to carry, in addition to voice, a diverse mix of information types. This has led to the development of a set of proposals for a telecommunication system, to carry these diverse information units in an economic fashion. The proposed telecommunication network is known as Broadband-ISDN (BISDN).

The development of this type of network has induced significant research effort. The principal subject of this thesis, is the development of control structures and techniques to improve the performance of BISDN networks. At the outset it should be noted that although this work focuses on BISDN, many of the conclusions are relevant to other high speed network technologies such as high-speed packet switched computer networks (e.g. IBM's plaNET) and Asynchronous Time Sharing (ATS) based broadband networks.

1.1 BISDN in perspective

BISDN [1] is the high bandwidth multimedia telecommunication network, proposed by CCITT^{#1}, to incorporate broadband features into the Integrated Services Digital Network (ISDN). It is tailored to become the universal **future** network, scheduled for implementation by telecommunication authorities worldwide within this decade. This network will need to handle a variety of types of service, with diverse demands on the network in terms of the bit rate and burstiness required. Continuous as well as variable bit rates will be serviced, e.g. data, voice, still and moving pictures, and multimedia applications (see for example tables 2.1 to 2.5 in [2] adapted from [3]). Asynchronous Transfer Mode (ATM)—a fusion between packet switching techniques and synchronous time division multiplexing [4]—will be the transfer mode for implementing BISDN, as already agreed by CCITT [1]. Information streams are divided into fixed length, self routing packets, commonly referred to as **cells**^{#2}, which are directed through the network by fast hardware switches. The bandwidth allocated to a connection may vary over the lifetime of the connection, hence ATM offers multiplexing and buffering within

^{#1} CCITT—International Telegraph and Telephone Consultative Committee—is the international body responsible for setting standards for public telecommunication.

^{#2} Cells are 53 bytes long, with 5 bytes assigned for the cell header and the remaining 48 bytes assigned for the information field of the cell.

the network to allow more effective use of the resources. However buffering and multiplexing leads to cell-delay and (possibly) cell-loss. At the same time, network users require guaranteed levels of performance.

These requirements, coupled with the wide range of traffic characteristics and quality of service constraints, as well as the geographic distribution and large dimension of the network, lead to some substantial problems in the control of BISDN networks.

BISDN has been extensively researched (especially during the past 2-3 years), as evidenced by the large body of published papers, see for example: the proceedings of INFOCOM, GLOBECOM, ICC, ITC, ATRS, ABSSS, to name but a few; the journals devoting whole issues to BISDN [5], [6], [7], [8]; the books published [2], [9], [10], [11]; and the large number of CCITT recommendations—a listing of CCITT recommendations for BISDN appears in [2]. Even with this large body of published works there are still substantial unresolved problems of control in BISDN networks. See, for example: the guest editorial comments in [12] (they state that: "The international telecommunications community **fully** appreciates the complexity of the issue and, to cope with this problem, proposed a large variety of congestion control techniques. Many researchers believe that there is no silver bullet and that control of high-speed packet networks can be obtained by executing several concurrent mechanisms..."); [13] the CCITT recommendations on traffic control and congestion control; Jain [14], and Boyer [15], for an enlightening discussion; and [16] the guest editorial comments of the JSAC special issue on congestion issues in BISDN for a brief discussion of some of the control difficulties (they state that: "the dynamic, heterogeneous, time-varying network environment, with different service requirements is a significant factor in the design of controls"; and that "the design of the entire system and the interaction of the various components is often more important than the optimisation of individual components").

1.2 Motivation for our approach

In this thesis, we focus on a structured approach for the solution of the complex control problem of BISDN discussed above. Initially we motivate an appropriate structure that can formally address complexity, and then offer specific solutions in the key generic **functions** areas of CCITT [13] – flow control, call admission control and **service-rate**^{#1} **control**^{#2}. We have been guided by the general objectives of traffic control and congestion control in BISDN as defined in [13]: "to protect the network and the user in order to achieve network performance objectives, with an additional role of network

^{#1} Cell-service-rate, bandwidth and capacity are equivalent terms in this context; for consistency we will use service-rate throughout this thesis.

^{#2} Control is used as the generic word for management, control, and allocation (see footnote on page 25 for a loose definition of these terms (as used by telecommunication theoreticians).

resource optimisation". However, we demonstrate that not only can high resource utilisation be obtained, but also very tight control of the network performance can be achieved by using appropriate problem formulations and solutions.

Additionally, as it is increasingly noted, communication networks must have satisfactory dynamic as well as steady-state performance (e.g. Van As [17], Tipper et al [18], Lovegrove et al [19], Bolot et al [20], the session on dynamic phenomena at GLOBECOM'92). Nonstationary conditions occur in communication networks when the statistics of the traffic arrival processes or queue service processes vary with time (for example due to nonstationary input loads, topological changes to the network or failures of network resources). This nonstationary behaviour is particularly significant in the context of BISDN networks because of the mix of traffic types and the nature of resource sharing. Therefore, the dynamic aspects of network behaviour **cannot** be ignored.

Moreover, we deviate **from** the norm and assume that the user is **not** able to declare all traffic characteristics of the offered traffic in advance, i.e. at the call setup. Therefore we do not base our control decisions on any user declared **parameters**^{#1} (apart from the peak rate of a special group of traffic, which we define as **controllable**^{#2}). Our assumption stems from the fact that in the majority of cases not all the characteristics of real-time traffic (as for example the mean value, the burstiness factor, etc) are known in advance. It is stated [15] that in most cases the peak rate is the only traffic parameter that the users are able to declare at the call set-up. Notwithstanding the above argument, even if the statistical parameters are reasonably well known, there are still substantial problems in using these traffic descriptors for network control in an open loop fashion. (For example enforcement **issues**^{#3}; also see [21] in which, using a real video sequence of 30 minute duration, they demonstrate that different sources with identical statistical characteristics can experience cell-loss rates that differ from each other by several orders of magnitude.) Also, on a more philosophical note, the flexibility of transmitting information on demand, in real time, is taken away from the user [15].

So, our control philosophy is principally guided by the following key assertions (elaborated at the relevant parts of the thesis):

- a structured approach is necessary due to the complexity of the system (see discussion in chapter 2).
- the dynamic aspects of network behaviour cannot be ignored.
- the use of formal control theory is worthy of investigation. We also allege that feedback, possibly supplemented with feedforward control, is an essential

^{#1} A standardised traffic descriptor is often discussed in the literature, in which a set of standard traffic parameters is available that **completely** characterises the behaviour of a traffic source.

^{#2} The definition of controllable and uncontrollable traffic appears in section 3.2.1, page 31

^{#3} References abound in section 3.1.1 page 26.

component for an effective and efficient control system (e.g. see discussion in section 2.2.3).

- there is no need for user declared parameters, except for the peak rate which is easily enforceable at the ATM layer Service Access Point (e.g. see discussion in section 3.1).

1.3 Overview of the thesis

The details of the organisation of the thesis are as follows:

The thesis is divided into six main chapters, and one appendix. The structuring of this work is slightly non-conventional. We have split the introduction between this chapter and the next, in order to highlight (and justify) our control philosophy. Additionally, we present the literature reviews and local appendices within the chapters that they best relate to. Therefore we present the table below as a guide mapping the "traditional" chapter classifications with those used here.

| | Introduction | Literature survey | Theory | Results and their discussion | Conclusion | Other |
|------------|--------------|-------------------|--------|------------------------------|------------|-------|
| Chapter 1 | ✓ | ✓ | | | | |
| Chapter 2 | ✓ | ✓ | ✓ | | | |
| Chapter 3 | | ✓ | ✓ | ✓ | | |
| Chapter 4 | | ✓ | ✓ | ✓ | | |
| Chapter 5 | | ✓ | ✓ | ✓ | | |
| Chapter 6 | | | | | ✓ | |
| Appendix A | | | | | | ✓ |

Table 1.1. Thesis guide

An overview of the rest of the chapters in the thesis follows.

In chapter 2 we show that a hierarchical organisation of tasks is necessary in a BISDN network due to the complexity of the system, that is, because of its large dimension and physical distribution in space with different event time scales ranging over several orders of magnitude. We propose a novel hierarchical structure based on the **system** behaviour in both time and space. We also argue that, for the effective and efficient control of BISDN: feedback, supplemented by feedforward control, must be considered; various

dynamic modelling techniques **are** essential; **and that** the **tradeoff** between service-rate, buffer-space, cell-loss and cell-delay cannot be ignored.

In chapter 3 we solve the flow control and call admission control problems together, thus taking into account the interaction between them. We aim to use a feedback control system to maintain the Quality of Service (**QoS**) close to a target value, irrespective of variations in traffic (the disturbance). We define two distinct groups of traffic (controllable and uncontrollable), which allows us to introduce the concept of network controllability. Controllability is achieved by simply bounding the uncontrollable traffic. Bounding of the uncontrollable traffic becomes the sole role of CAC. The feedback signal is derived **from** a network performance monitor (which predicts the p th percentile of the buffer distribution). The controller regulates **QoS** by manipulating the flow of controllable traffic into the network. Controllability guarantees that the network can be operated efficiently (theoretically at 100% utilisation) and still provide the user with tightly regulated **QoS** (set at any desired target value). We employ the general methodology of adaptive control (featuring both adaptive feedback and adaptive feedforward) to solve the difficult problems of CAC and flow control together. By using this approach we are able to do away with the restrictive requirements of existing schemes (for example the need to declare a complex traffic descriptor at the call connection request). The performance of the derived algorithm is illustrated via analysis and simulation. The presented solution is in a form suitable for incorporation in a hierarchically organised control structure.

In Chapter 4 we focus on service-rate control at one level only—that of the virtual path (VP) level. We present a novel scheme that dynamically allocates service-rate by using the state of the buffers in the network as a feedback signal. In particular, we use a dynamic fluid flow type equation to model the VP and formulate precise problems for the control of service-rate. The interactions within the nodes spanned by a VP, as well as the interactions between the VPs sharing a link, are addressed in the problem formulation. We investigate the use of optimal single level and multilevel control theory concepts (featuring decomposition and coordination) in the solution of the service-rate control problem. The performance of the scheme is investigated using simulation. The form of the solution is suitable for incorporation in a broader, hierarchically organised, control structure. We also use an **ad hoc** approach to extend published results on the coordinated decentralised control of large scale nonlinear systems to deal with **time-**delayed systems.

Chapter 5 presents an illustrative example of a hierarchically organised control structure (that can form a part of the overall solution) for the control of service-rate. At the highest level, an extension to an existing service-rate allocation algorithm is presented and its integration with lower levels discussed. At the lowest level, a novel link service protocol (based on heuristic arguments) is described. For the **intermediate** levels we make use of the solution of the service-rate at the VP level described in Chapter 4. To illustrate the flexibility of the hierarchically organised control structure, we formulate another intermediate level and discuss its integration with the overall structure.

Finally, the conclusions appear in Chapter 6 together with a listing of the specific contributions of the thesis and the suggestions for **further** work in this area.